Cluster Analysis: Past, Present and Future

Brian S. Everitt

Cluster analysis is a generic term for a wide range of numerical methods for examining data with a view to detecting, uncovering or discovering groups or 'clusters' of objects or individuals that are (1) homogeneous and (2) separate. Many of the clustering methods is use today have resulted from a crossfertilization between disciplines such as psychology, biology and psychiatry on the one hand and mathematics and statistics on the other. The result has been a considerable amount of ad hoc development and 'reinvention'—as somebody once remarked 'there may be as many clustering techniques as there are cluster analysis users'. In the last five years or so, cluster analysis has become part of the data mining industry and has found application in grouping together genes with similar patterns of expression.

This talk will look at some early papers on cluster analysis that were important to the speaker, the current state of the methodology and speculate briefly about its future development.